

# A stochastic approach for extending the dimensionality of observed datasets

*Tamás Várnai*  
*Joint Center for Earth Systems Technology,*  
*University of Maryland Baltimore County*

Submitted to Geophysical Research Letters

## Popular Summary

This paper addresses the problem that in many cases, observations cannot provide complete fields of the measured quantities, because they yield data only along a single cross-section through the examined fields. One such example is in-situ airplane measurements of cloud properties along the airplane's flight path. Other examples are vertically-pointing ground-based instruments that measure cloud properties only above their fixed location: It is true that as the wind blows clouds over these ground-based instruments, one can use the time series of measurements to reconstruct cloud structure along the line of wind direction—but cloud properties outside this line remain unknown. The paper describes a new Fourier-adjustment technique that allows existing fractal models to build realistic surroundings to the measured cross-sections. This new approach allows more representative calculations of cloud radiative processes and may be used in other areas as well.

# **A stochastic approach for extending the dimensionality of observed datasets**

**Tamás Várnai**

Joint Center for Earth Systems Technology,  
University of Maryland Baltimore County,  
Baltimore, Maryland

*Prepared for Geophysical Research Letters*

September 2002

Author address: Code 913, NASA GSFC, Greenbelt, MD 20771.  
e-mail: varnai@climate.gsfc.nasa.gov

## **Abstract**

This paper addresses the problem that in many cases, observations cannot provide complete fields of the measured quantities, because they yield data only along a single cross-section through the examined fields. The paper describes a new Fourier-adjustment technique that allows existing fractal models to build realistic surroundings to the measured cross-sections. This new approach allows more representative calculations of cloud radiative processes and may be used in other areas as well.

## Introduction

In many cases, observations cannot yield the full spatial distribution of geophysical parameters, because they provide data only along a single cross-section of the complete field. One such example is in-situ airplane measurements of cloud water content along the airplane's flight path. Other examples are vertically-pointing ground-based instruments that measure cloud properties only above their fixed location—for example, Millimeter Wavelength Cloud Radars and Microwave Radiometers, which measure the altitude-dependent cloud water content and the vertically integrated cloud water path, respectively. It is true that as the wind blows clouds over these ground-based instruments, one can use the time series of measurements to reconstruct cloud structure along the line of wind direction—but cloud properties outside this line remain unknown.

Unfortunately, knowledge of a single cross-section is not sufficient for some types of studies, such as studying radiative processes in heterogeneous clouds. So far, researchers have had two options for dealing with this situation. The first possibility is to assume no variability along the missing dimension and to perform 2D radiative calculations instead of 3D ones (e.g., Hignett and Taylor 1996, Zuidema and Evans 1998, Los and Duynkerke 2000). The inherent limitation of this approach is, of course, that 2D calculations cannot capture the full complexity of 3D processes (e.g., Barker 1996). The second possibility is to use observations for calculating the measured field's statistical parameters, then to use a stochastic model to create complete fields with the desired parameters, and finally, to perform 3D radiative calculations for the created fields (e.g., Barker and Davies 1992, Davis et al. 1997, Iwabuchi and Hayasaka 2002). The main

limitation of this approach is that although the statistical parameters reproduced by stochastic models strongly influence the cloud fields' radiative parameters, they do not determine them completely.

This paper offers a third choice to researchers: It proposes a Fourier-adjustment technique that allows stochastic models not only to reproduce the observations in a statistical sense, but also to exactly fit the available observations. As a result, the new approach can provide realistic surroundings to the actual measurements, thus extending the observations' dimensionality. Because the stochastic extension process creates fractal structures, the procedure may be termed "fractal extrapolation".

In the area of cloud radiative transfer, the main advantage of fitting stochastic models to the measurements is that—at least near the observed cross-sections—this allows the models to reproduce even the cloud properties not considered explicitly in the model. As a result, the new approach generates cloud structures that will allow more representative radiative transfer calculations than either conventional stochastic models or stand-alone observations would.

If additional information is available for areas outside the measured line, the approach can also use this data to provide an integrated, holistic view of clouds. For example, the new method can combine the ground-based information in  $x$  and  $z$  directions with satellite information about variations in the  $y$  direction and thus provide an integrated 3D view of clouds.

This paper first describes the proposed approach for a relatively simple case, and then it discusses how this basic method can be extended to handle increasingly complex situations.

## Description of Basic Methodology

To present the basic concept of the new approach, let us assume that on an overcast day, we measure the  $W(t)$  time series of vertically integrated water path values using a vertically-pointing, ground-based Microwave Radiometer (MWR). Adopting the widely used frozen turbulence assumption and knowing the wind speed and direction, we can convert the time series of measurements into a 1D cross-section of the cloud field,  $W(x)$ . Our goal is then to extend the data into the missing cross-wind direction, thus obtaining a realistic  $W(x,y)$  field.

The proposed approach works for any stochastic model that is based on the following three-step procedure:

Step 1: Generate a suitable set of random Fourier coefficients.

Step 2: Do an inverse Fourier transform to get an initial 2D field.

Step 3: Apply a transformation  $T$  to get the final  $W(x,y)$  field.

In Step 1, most current fractal models generate the initial Fourier coefficients as a complex Gaussian white noise (0 mean, unit standard deviation, and uniform phase-distribution between 0 and  $2\pi$ ), and the coefficients are multiplied by an appropriate factor to achieve the desired scaling properties. As for Step 3, the proposed methodology works with any  $T$  transformations used in current stochastic models, such as addition and multiplication in Fractional Brownian Motion (e.g., Barker and Davies 1992), an extra exponentiation in Frank Evans's model (Evans 1993), and a final low-pass filtering in fractional integration models (Schertzer and Lovejoy, 1987).

The proposed approach considers that the Fourier transform decomposes the actual data into a sum of sine waves in both the 1D and 2D situations:

$$W(x) = \frac{1}{N_x} \sum_{k=0}^{N_x-1} H_k \cdot e^{\frac{-i2\pi xk}{N_x}} \text{ and} \quad (1a)$$

$$W(x,y) = \frac{1}{N_x} \sum_{k=0}^{N_x-1} \left( \frac{1}{N_y} \sum_{l=0}^{N_y-1} H_{k,l} \cdot e^{\frac{-i2\pi y l}{N_y}} \right) \cdot e^{\frac{-i2\pi x k}{N_x}}, \quad (1b)$$

where  $N_x$  and  $N_y$  are the number of data points along the  $x$  and  $y$  axes, respectively ( $0 \leq x \leq N_x-1$ , and  $0 \leq y \leq N_y-1$ ).  $k$  and  $l$  are the wavenumbers along these axes, and the complex Fourier coefficients  $H_k$  and  $H_{k,l}$  describe the amplitude and phase of each sine wave. For simplicity, we assume that observations are available along the  $y = 0$  line. Because the generated fields will be periodic ( $W(x,0) = W(x,N_y-1)$ ), we will obtain surroundings on both sides of the observed line. For visual clarity, we may then use a cyclic shifting of the generated field to place the observed line into the center.

Since the measured 1D data sequence  $W(x)$  represents a cross-section of the 2D field  $W(x,y)$ , we know that the 1D waves we obtain from the Fourier analysis of  $W(x)$  really arise from the combination of the 2D waves that describe the  $W(x,y)$  field. Thus the key task of the new method is to generate a set of 2D waves that, when combined, form the 1D waves we obtained from the 1D data series. This requirement can be formulated precisely as follows: The comparison of Eqs. (1a) and (1b) reveals that in order to get  $W(x) = W(x,0)$ , we need to satisfy the equation

$$H_k = \frac{1}{N_y} \sum_{l=0}^{N_y-1} H_{k,l}. \quad (2)$$

To satisfy this equation, we need to adjust the initial 2D waves generated randomly in Step 1. For the adjustment, we first need to calculate the “desired” 1D waves ( $H_k$  values)

by applying to our 1D data series the inverse of transformation  $T$  and then performing a forward Fourier transform. Then we can adjust the initial, random 2D waves ( $H_{k,l}^{init}$  values) according to the equation

$$H_{k,l}^{new} = E_{k,l} H_{k,l}^{init} \frac{H_k}{\frac{1}{N_y} \sum_{l=0}^{N_y-1} E_{k,l} H_{k,l}^{init}}. \quad (3)$$

In this equation,  $E_{k,l}$  is a complex number used to ensure that the adjusted 2D waves combine with an appropriate effectiveness along the line of measurement (i.e., that they tend to cancel out or strengthen each other to an appropriate degree). The  $E_{k,l}$  values are calculated in four steps. First, we fit a smooth function  $G_k$  on the  $|H_k|$  values around  $k$  (for example in the form  $G_k = a k^b$ , using the empirical  $a$  and  $b$  values that ensure a best fit).

Second, we calculate the  $C_k = \frac{|H_k|}{G_k}$  values to estimate how effectively the real  $W(x,y)$

field's 2D waves combine along the line of measurements: whether they tend to strengthen or cancel out each other. Third, we calculate the  $D_k(y)$  values that describe how effectively our randomly generated 2D waves with wavenumber  $k$  combine along the lines at each  $y$  value:

$$D_k(y) = \frac{\left| \frac{1}{N_y} \sum_{l=0}^{N_y-1} H_{k,l} \cdot e^{\frac{-i2\pi y l}{N_y}} \right|}{\frac{1}{N_t} \sum_{t=1}^{N_t} \left( \frac{1}{N_y} \sum_{y=0}^{N_y-1} \left| \frac{1}{N_y} \sum_{l=0}^{N_y-1} H_{k,l} \cdot e^{\frac{-i2\pi y l}{N_y}} \right| \right)}. \quad (4)$$

In this equation, the numerator is the amplitude of the 1D wave that results from the combination of 2D waves in our case, and the denominator is the mean value of amplitudes that could be expected from the combination of randomly chosen 2D waves.

In the denominator,  $N_t$  is the number of test arrangements that are used to calculate the mean expected amplitude (usually less than 10 is sufficient), and  $C_{k,l,t}$  is a random complex number with unit length. ( $C_{k,l,t}$  randomly changes the phase of 2D sine waves so that in each arrangement, they combine with a different effectiveness.) Fourth, we select for each  $k$  value a corresponding  $y_k$  such that  $C_k \approx D_k(y_k)$ , and then we get the  $E_{k,l}$  values as

$$E_{k,l} = e^{\frac{i2\pi y_k l}{N_y}} \quad (5)$$

Let us note that if, because of the random arrangement of our initial 2D waves, no suitable  $y_k$  value is found for which  $C_k \approx D_k(y_k)$ , we need to generate another set of random  $H_{k,l}^{init}$  values for this particular  $k$  and try the adjustment procedure again.

Because the adjustment of  $H_{k,l}^{init}$  in Eq. (3) means a multiplication by a random factor, it does not spoil the statistical properties of the initial 2D waves, and so it does not change the statistics of the final  $W(x,y)$  field. An initial test result from this procedure is shown in Figure 1.

## Discussion of Possible Extensions to More Complex Situations

### *Broken Cloud Fields*

In overcast situations, the method described above generates  $W(x,y)$  fields that can be used in 3D radiative transfer studies much the same way as the  $W(x,y)$  fields derived from satellite images have been used in the past (e.g., Barker and Liu, 1995, Liou and Rao 1996, O'Hirok and Gauthier 1998). In the case of broken clouds, however, a small complication arises: Transformation **T** in Step 3 of the basic stochastic model procedure involves a cut that sets all negative  $W$  values to zero in the cloud-free areas (e.g., Barker

and Davies 1992). This implies that the inverse transformation  $T^{-1}$  in Step 1 (needed to get the desired  $H_k$  coefficients) must replace the zero values in the observations by stochastically generated negative values, for example using the midpoint displacement method. A sample result for broken clouds is shown in Fig. 2.

### *Extrapolation of 2D Data Into a 3D Field*

The proposed approach can also be extended to generate 3D cloud structures based on measured vertical ( $x$ - $z$ ) cross-sections of cloud properties. For homogeneous turbulence, the 2D to 3D extrapolation is directly analogous to the 1D to 2D procedure described above, and it needs only straightforward changes to accommodate the additional dimension. (For example,  $H_{k,l}$  should be replaced by  $H_{k,l,m}$ .) Real clouds, however, often display systematic altitude-dependent features, and in such cases it may be more appropriate to adopt a layer-by-layer approach in which the 1D data series for a given altitude is used to generate a 2D field for that altitude. In this approach we need to pay special attention to preserving throughout the scene the oblique and vertical correlations that are present in the measured ( $x$ - $z$ ) plane. To achieve this, we should not generate brand-new random 2D sine waves ( $H_{k,l}^{init}$  values) for each altitude, but instead we should only adjust the 2D sine waves from one level to the next by an appropriate amount.

### *Anisotropic Cloud Fields*

The proposed approach can also be extended to situations when we have additional information about variability outside the measured cross-section. For example,

a whole-sky imager or a satellite image may give statistical information on the clouds' horizontal anisotropy—e.g., whether cloud streets occur. Any observed anisotropy can be reproduced if the initial  $H_{k,l}^{init}$  coefficients are multiplied by an appropriate anisotropy-function  $f(k,l)$ , because the subsequent adjustment by Eq. (3) does not change the field's anisotropy ( $H_{k,l}^{new}$  will be just as anisotropic as  $H_{k,l}^{init}$ .) While future studies are needed to determine the most appropriate function for describing horizontal cloud anisotropy, the following general form may be widely applicable:

$$f(k,l) = 1 + g(k) \cdot h\left(\frac{k}{l}\right), \quad (6)$$

where  $g$  describes how the magnitude of anisotropy changes with scale (e.g., small-scale turbulence is fairly isotropic even in cloud streets), and  $h$  describes the angular shape of anisotropy.

### *Integration of Observed Vertical Cross-sections With Images Depicting Horizontal Variability*

The proposed approach may reproduce the main features of actual clouds even outside an observed vertical cross-section, if the data can be co-located with images that depict the cloud field's horizontal structure. This can be the case if, for example, a vertically pointing lidar and an imager are mounted on the same airplane, or if a satellite passes over the location of our ground-based instruments. In such cases the model could combine the ( $x$ - $z$ ) cross-section with ( $x$ - $y$ ) images to give us an integrated, 3D view of clouds. For this, Step 1 should not generate the initial 2D sine waves ( $H_{k,l}^{init}$  values) randomly; instead, it should take them from a 2D Fourier analysis of the available image.

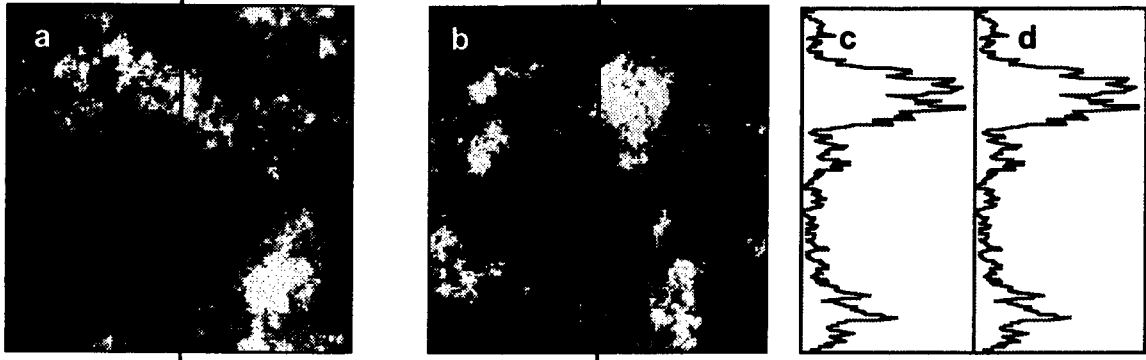
Let us note, however, that since the satellite information will usually disagree slightly with the ( $x$ - $z$ ) measurements, the model-generated fields will not match the observed ( $x$ - $y$ ) fields exactly; but they are expected to reproduce the dominant features of the actual clouds, for example, the location of gaps and turrets.

**Acknowledgments** Funding for this research from the NASA EOS Project Science Office (under Grant NAG5-6675) was greatly appreciated. We are also grateful to Laura Atwood for proofreading the manuscript.

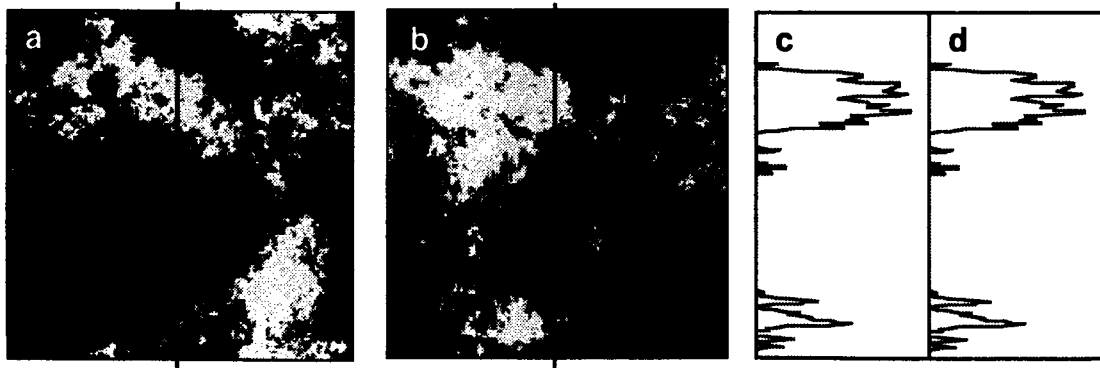
## References

- Barker, H. W., 1996: Estimating cloud field albedo using one-dimensional series of optical depth. *J. Atmos. Sci.*, **53**, 2826–2837.
- Barker, H. W., and D. Liu, 1995: Inferring optical depth of broken clouds from Landsat data. *J. Climate*, **8**, 2620–2630.
- Barker, H. W., and J. A. Davies, 1992: Solar radiative fluxes for stochastic, scale-invariant broken cloud fields. *J. Atmos. Sci.*, **49**, 1115–1126.
- Davis, A., A. Marshak, R. Cahalan, and W. Wiscombe, 1997: The Landsat scale break in stratocumulus as a three-dimensional radiative transfer effect: Implications for cloud remote sensing. *J. Atmos. Sci.*, **54**, 241–260.
- Evans, K. F., 1993: Two-dimensional radiative transfer in cloudy atmospheres: The spherical harmonic spatial grid method. *J. Atmos. Sci.*, **50**, 3111–3124.

- Hignett, P., and J. P. Taylor, 1996: The radiative properties of inhomogeneous boundary layer cloud: Observations and modelling. *Q. J. R. Meteor. Soc.*, **122**, 1341–1364.
- Iwabuchi, H., and T. Hayasaka, 2002: Effects of horizontal inhomogeneity on the optical thickness retrieved from moderate-resolution satellite data. *J. Atmos. Sci.*, **59**, 2227–2242.
- Liou, K. N., and N. Rao, 1996: Radiative transfer in cirrus clouds. Part IV: On cloud geometry, inhomogeneity, and absorption. *J. Atmos. Sci.*, **53**, 3046–3065.
- Los, A., and P. G. Duynkerke, 2000: Microphysical and radiative properties of inhomogeneous stratocumulus: Observations and model simulations. *Q. J. R. Meteor. Soc.*, **126**, 3287–3307.
- O'Hirok, W., and C. Gauthier, 1998: A Three-Dimensional Radiative Transfer Model to Investigate the Solar Radiation within a Cloudy Atmosphere. Part I: Spatial Effects. *J. Atmos. Sci.*, **55**, 2162–2179.
- Schertzer, D., and S. Lovejoy, 1987: Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.*, **92**, 9693–9714.
- Zuidema P. and K. F. Evans, 1998. On the validity of the Independent Pixel Approximation for the boundary layer clouds observed during ASTEX. *J. Geophys. Res.*, **103**, 6059–6074.



**Figure 1.** (a) A cloud field generated by the stochastic model of Frank Evans (Evans 1993). The black line indicates the position of hypothetical cloud observations  $W(x)$ . (b) A sample cloud field generated by the proposed method that fits the hypothetical observations. (c) and (d) The two fields' cross-sections along the specified line. (The two cross-sections match exactly, and the two fields have very similar statistical properties, namely histogram, anisotropy, scaling, and intermittency.)



**Figure 2.** Same as Figure 1, but for broken clouds.